

---

# The Archival View of Technology: Resources for the Scholar of the Future

DAVID S. ZEIDBERG

---

## ABSTRACT

WHERE EARLY MANUSCRIPTS AND PRINTED BOOKS have survived because they are made of durable materials, electronic records are vulnerable to alteration and data loss. Scholars can trace most traditional materials to the source of their creation through various versions and editions of a work but do not enjoy the same research paths with electronic records. Archivists need to apply traditional principles of records management to electronic documents if these records are to survive for research by future scholars. Additionally, archivists must ensure that the electronic records preserved remain accessible as hardware and software change over time.

## TRADITIONAL RESOURCES AND TRADITIONAL SCHOLARSHIP

Since the invention of writing, human beings have not only recorded information and ideas they thought important but have attempted to save that information as well. Early documentation was mostly practical, but developed toward historical perspective—records of harvests; the exchange of goods; the lineage of a family; the chronicle of a monarchy, society, or culture. Literary records, begun in an oral-formulaic tradition, came later. Throughout the history of the written word, people serving in roles equivalent to our modern concept of librarians and archivists have attempted to preserve, arrange, and describe these original documents, not only to save the ideas they contain but also to keep some sense of the process of creating those ideas. Scholars seek out these documents, even if they have been published, to see them “first-hand” and to understand how they were

created. Original documents hold physical evidence that transcends the ideas that the words, sentences, and paragraphs contain.

The best scholarship in the humanities still emanates from documentary research. Textual transmission plays a key role in determining the accuracy of a resource, and corruption from one transmission to the next—as manuscripts are copied, as different editions are set in type, as editors make unfounded decisions—can produce texts far from the author's original intentions. One example of textual corruption concerns the printing of Archimedes's works. The first two editions appeared in Venice and Basel in 1543 and 1544 respectively. The printers based their texts on manuscripts available to them. These manuscripts, it turned out, were far from the texts Archimedes is thought to have left when he died in 212 B.C. Federico Commandino, an extraordinary Renaissance scholar of physics and mathematics, found the printed texts troubling and set about to produce a new edition around 1550.

Commandino accomplished the textual restoration through his understanding of classical Greek, and, more importantly, through his grasp of the process by which the texts had been transmitted from one manuscript to the next from Archimedes's time through the Roman and Byzantine periods to the Renaissance, and with translation into Latin as well along the way. Beginning in the Hellenistic period, scholars would add glosses of difficult words in the text, called *lemmas*, along with commentaries, called *scholia* (see Grafton, 1997, p. 157ff., for an explanation of this process and its effect upon original texts). Commandino was able to work back through the *scholia* and *lemmas* of the two contemporary printed editions and earlier manuscripts, making corrections and, in some cases, eliminating erroneous glosses altogether. Paolo Manuzio printed and published the results of Commandino's restoration of the Archimedes text in 1558, and that edition remained definitive—the one from which all subsequent editions were published—through the nineteenth century.

Sound textual scholarship produces definitive editions, and sometimes the impact of those editions can change the world. Perhaps the most far-reaching example in Western culture occurred in the first half of the sixteenth century during the Reformation. Martin Luther's 1517 publication of his ninety-five theses against the sale of indulgences may be the best remembered "document" of the Reformation, but biblical scholarship both within and outside the church probably had more long-lasting effect on the movement. Humanist scholars, such as Luther, Desiderius Erasmus, and William Tyndale worked on new translations of the Bible while Cardinal Ximenes worked a more subtly presented new translation into a new polyglot Bible in Alcala, Spain. At issue was the accuracy of the existing Latin Vulgate, Jerome's fourth century translation from the original Hebrew and Greek, which had remained the official Catholic version for a thousand years. David Daniell (1994) writes: "Though limited and

in places misleading and inaccurate, it was powerfully defended; attempts to restore knowledge of the texts, the Greek of the new Testament and the Hebrew of the Old, were usually branded as heresy" (p. 4).

Erasmus led the way in 1516 with his *Novum instrumentum*, a new Latin translation of the New Testament printed in parallel to the original Greek text. While Erasmus's intention was to "correct the Vulgate" as Daniell observes, having his translation printed next to the Greek from which it was made gave a more scholarly than political tone to the product. Luther's 1522 German translation and Tyndale's English version of 1526 were more polemically driven editions, made in defiance of the church, that aimed at putting accurate vernacular versions of the Bible in the hands of lay people. Indeed, for his efforts, Tyndale was eventually captured in 1535, tried and imprisoned, and burned as a heretic in 1536.

Within the Catholic church, at the time Erasmus was producing his New Testament translation, Cardinal Ximenes was directing a team of scholars to produce the Complutensian Polyglot Bible, which eventually appeared in 1522. Across a single page, one could now find the original Old Testament Hebrew, the Greek translation of the Hebrew, Jerome's Vulgate, Aramaic commentaries on the Hebrew and, perhaps most important, a new Latin translation. Like Erasmus, the intent was to present the latest textual scholarship that "allowed the Vulgate to be challenged, and advanced understanding of the original texts" (Daniell, 1994, p. 10). In all these examples, the fact that these scholars could retrace the steps by which the source documents were created afforded them the opportunity to make more sound interpretations of what they read.

This is not to say that scholars always interpret what they read correctly. Misinterpretation of original resources can create wrong conclusions that can be perpetuated for generations. As one final example of traditional textual scholarship, Anthony Hobson notes that the misreading of some early sixteenth century letters by nineteenth century scholars caused them to conclude that Aldus Manutius had a bindery attached to his printing shop (Hobson, 1998, pp. 237-45). The misinterpretation was not as questionable as the perpetuation of the myth for more than 100 years by generations of scholars who took the first mistake as truth. The larger point again is that the error was traceable and could be corrected. Will future scholars have the same opportunity working with today's information when documents are created electronically; when systems for authenticating, organizing, and preserving this new archive of information are in only fledgling states; and when archivists are thwarted by public indifference to the authority of the text?

### THE CREATION OF DIGITAL "MANUSCRIPTS" AND ELECTRONIC PUBLISHING

Early manuscripts, books, maps, drawings, and other three-

dimensional objects of research value have survived because they are durable. Millions of these materials exist in libraries and archives worldwide. In their short lifetime, electronic records are far more numerous already but at much greater risk of loss. Furthermore, as electronic publishing proliferates, we are witnessing difficulties with electronic texts not unlike those faced by the first European printers in the latter half of the fifteenth century. Then printers were limited by the manuscripts known and available and by their accessibility to humanist scholars who could help with the editing of the first printed editions. Today, the discerning reader who looks at an electronic publication must question how the publication came into being: Is what she or he is reading what the author truly wrote? How was it refereed? How was it edited and by whom? Has it been protected through encryption from unauthorized alteration?

If it is the electronic version of a previously published work, from which edition was it transmitted into electronic form? Some electronic publishing entrepreneurs, in their rush to get popular texts into electronic format, have avoided copyright issues by only publishing works or editions in the public domain with no regard for textual accuracy or authority. Many of these electronic editions are of little use to scholars because they are either inaccurate or poorly edited versions of the texts (or both). Like the first printings of Archimedes in the sixteenth century described above, the publishing of some texts in electronic format has been limited by those available to the publisher, in this case those in the public domain to avoid copyright; like the 1543 and 1544 Archimedes editions, these modern texts are not necessarily the best texts available.

This is not to say that every electronic text is suspect. Many worthwhile bibliographic projects may be found on the Internet, where scholars are participating in constructing the content and are discerning in the editorial information they provide. A good example is the William Blake project based at the University of Virginia (<http://www.iath.virginia.edu/blake>) with contributions of original editions from major libraries and management by a team of academic scholars. Their work is scrupulous, but how is an undergraduate with a Blake assignment, for example, to choose between this site and an amateurish one, also to be found on the Internet, which has mounted Blake texts without permission, authorization, or editorial competence?

When archives and special collections libraries work with scholars to produce network surrogates of their original holdings, there can be educational benefits through the sharing of two-dimensional forms of original materials that some scholars might not otherwise have a chance to see. Certainly the content can be shared, as it has been in the past through microform and photographic copies. The digital versions of similar original works held at different libraries offer the further advantage that they

can be compared side by side. Sometimes one is also able to read a digitized surrogate more easily than an original because the photography can "bring back" erasures and palimpsests.

Of further issue to literary and historical scholars is the creative process itself. One has to ask how many of today's writers preserve one draft or version of a word-processed manuscript to the next so that scholars can understand the writer's thought processes. An original historical manuscript illustrates this point: in 1782, George III grudgingly capitulated to America's independence, more than a year after Cornwallis's surrender at Yorktown. In a manuscript preserved at the Huntington Library, George III writes: "Parliament having to my astonishment come into the ideas of granting ~~Independence~~ a Separation to North America, has disabled Me from longer defending the just rights of this Kingdom . . ." (George III, autograph letter . . . , 1782). As illustrated in this transcription, the King has crossed out the word "Independence" and written over it "a Separation." He cannot bring himself to utter or write the word "Independence" regarding the American colonies; he finds it too objectionable. Were George III to have had access to a word processor for composing his letters and documents, would we have ever seen this change or have had the opportunity to interpret the feelings behind the words? We probably would only see the final version, if even that were preserved.

Indeed, which electronic records to preserve is also an issue. As organizations turn to electronic record keeping, the archival principles behind a records retention schedule become even more important. At this writing in the fall of 1998, for example, Congress has given the United States Archivist an extension to develop a workable comprehensive plan to direct government agencies on which electronic records are important to save. This is not only a matter of what to store and what to delete, but also how to store it in a way that will eventually allow an archivist to read the files as she or he appraises, arranges, and describes them at the point they are processed and made accessible for research. With paper records, years can go by before the processing is done, but the files are still readable when they are addressed. With electronic records, the files will have to be migrated as new hardware and software are developed if the readability is to keep up with technological advances.

So, if all of the foregoing questions regarding the selection, authenticity, and accuracy of an electronic manuscript or text were properly resolved, *how* is today's archivist to preserve electronic documents for future scholars' use and assure that they will be accessible? With hardware and software changing and upgrading almost by the minute, how will a scholar in 2099 read an electronic manuscript written in the 1980s on a Macintosh with a Mac platform version of Microsoft Word 2.0? Ninety-nine point nine percent of PC-based users *today* could not open and read that document on their present personal computers.

## ARCHIVAL RESEARCH IN THE FUTURE

Archivists already face the daunting task of devising methods to organize and preserve electronic records and to assure their readability for generations to come. Each of these issues is enormously complex. In his article "New Roles for Special Collections on the Network," Peter Graham (1998) points out that whether materials deemed worthy of preserving are in the traditional formats of manuscript or print or in electronic format, "[t]he fact remains that for information to be available for any meaningful length of time, someone has to select it and take responsibility for it, which has been—and remains—the role of the library" (pp. 234-35). The standard archival methods of arrangement and description of documents and records can apply to electronic archives. There is an added advantage of keyword searching through a database that can take a researcher more quickly to the information she or he is seeking. If one were looking for correspondence of a particular person in a large archive, a keyword search of the name would take the reader to all the places in the archive where the person appears.

The keyword search provides an advantage, but it is not a substitute for other standard research strategies. If librarians and archivists were to forego their traditional methods of arrangement and description, thinking the expediency of keyword searching in a database would alone serve researchers, there would be severe losses in the understanding of information and knowledge. Thomas Mann (1993), for example, identifies eight "avenues" of access to library (and archival) resources:

1. controlled-vocabulary subject heading sources;
2. classified array of subject-grouped printed full texts [i.e., collections arranged on the shelves];
3. printed keyword indexes (which have substantial coverage not in computer formats);
4. printed citation indexes (which also have substantial coverage not found in computer databases);
5. published bibliographies (again, providing wide-ranging and deep coverage not duplicated by computers);
6. computer sources beyond those in avenues 1 and 7 (including CD-ROM, dial-up, or in-house; also encompassing bibliographic citation, full-text, network, and bulletin board forms—the whole range);
7. related record CD-ROMs;
8. people sources. (pp. 184-85)

These avenues of access, as written, apply to book collections. Archival arrangement and description practices parallel them, and the indexes and bibliographies can apply to archival holdings if they are included in them. Mann's point is that, if any of the avenues are neglected or missed, the researcher may miss complete comprehension of the resources available:

If any subject within the full circle of knowledge is not taught . . . , the void it leaves will tend to be filled in, within students' perceptions, by the other disciplines—with less than satisfactory results. . . . If vocabulary-controlled searching isn't taught, then people will mistakenly perceive keyword searching as "covering" that void. If the use of published bibliographies isn't taught separately, then people will mistakenly perceive computer searching as filling that need. If talking to people isn't emphasized, then students will try to overinflate the use of print or electronic source to try to cover that lack. The point is this: people will generally not allow themselves to perceive a gap in their knowledge; what they will do instead is to inflate the part they do grasp to take the place of the whole that they do not see. And if they get any results at all from the part, they will then "satisfice" with the results. Furthermore, they will mistakenly conclude that they have tried "everything" when in fact they have exhausted only the few avenues they do perceive, all the while missing much more than they find, but not being aware of it because they have indeed searched "everything" in the knowledge universe *as they perceive it*. A Methods model, more than any other, would correct this problem for researchers. It would give them the best map of the *whole* of the research universe that ought to be available to them. (pp. 182-83)

What is worrisome to most teachers and librarians is that the Internet and its resources are being sold as a fast track to information when collection builders know that what is actually on the Web is at the dictionary and encyclopedia level at best, with a few exceptions, such as the Blake project noted above. The risk is that, if untrained and uninformed researchers do not find what they are looking for on the Internet, they will not employ the other search strategies Mann has defined, and therefore neither find what they are researching nor perhaps even think it exists. If librarians and archivists are to provide comprehensive access to their resources, they must continue to use traditional methods of arrangement and description to provide the paths.

The other important issue is the ability to deliver the materials. For traditional resources, whose value as artifacts is as important as their content, it has always been a matter of physically preserving the books, manuscripts, prints, photographs, maps, and ephemera according to archival standards, housing the materials in a secure place, and providing organized access to them in a controlled environment that will protect the materials from damage and theft. As every library or archives administrator knows, these are costly functions but necessary to assure that original resource materials will be available from one generation to the next. Electronic archives can be preserved in more economical and various forms of storage—floppy disks and hard drives, CDs, tape, zip- and jazz-drives, bubble memory, and so on. Electronic archives can be copied relatively inexpensively as long as the technological platform remains the same.

The risks attendant to electronic archives are: (1) the protection of

the information against alteration, and (2) the change in technology platforms that may render the archive unreadable after only a few years. Methods of encryption and electronic "watermarking" can help with the former issue. These are not without their own expense, and library administrators eager to save money by abandoning preservation of traditional resources to embrace the digital format should be mindful that all preservation has a price. Peter Graham (1998) notes that the "greatest asset of electronic information is also its greatest liability . . . . And at all times in the electronic environment, the integrity or authenticity of the object needs to be guaranteed for the user to have assurance that the information is what it is expected to be (intellectual preservation)" (p. 234). Librarians and archivists guard against the defacement and alteration of artifactual materials by permitting them to be used only under observation. This has not prevented some from vandalizing materials or acting as self-appointed censors by tearing out pages or blackening texts found offensive to them. However, one can readily see the evidence of these catastrophes; the damage is obvious. Short of excised and destroyed pages, some damage may be reversible under current conservation practices.

Librarians and archivists will have to take special measures to prevent similar damage or alteration to electronic archives, especially when the change may not be readily apparent: passages offensive to the censor could be deleted without an unsuspecting reader even realizing this later. Worse, text could be replaced with the censor's own political viewpoint, misleading the reader even further. Encryption producing "read-only" access will be a deterrent, although determined hackers might decode the encryption in time. Electronic watermarking, hidden in the electronic text, could help to assure authenticity, although this process too can be defeated. Neither is an inexpensive process.

Delivering electronic archives written with now-obsolete hardware and software poses a greater challenge. Will libraries have to become museums of equipment and software held in the chance that a reader might require access to a document composed on a specific machine with a specific version of software? There are earlier forms of electronic recording and compilation already in repositories with no means to play or read them. For example, the forerunner of the audiotape recorder was a wire spool machine. Some archives have examples of these recordings with no working equipment to play them. Perhaps more relevant examples are institutional archives whose organizations made use of early punch card computing and mainframe tapes to store institutional data and information, again with no functional equipment to provide access to this material today.

The proliferation of the personal computer less than thirty years ago put the creation of electronic manuscripts and databases in the hands of the individual. Memory, speed, and storage capacities have made quan-

tum leaps, and the simplest word-processing programs have given way to complete desk-top publishing. Most people involved in the technological development have paid little attention to what has come before, so that one is lucky to be able to read a file on an earlier version of the same software one is running presently. Cross-platform access is still almost unheard of for the average PC user. So if archivists are to preserve and provide access to electronic archives, will they also have to acquire and maintain equipment and software on which each collection was created?

Fortunately, there are some people looking for alternatives to this overwhelming and bleak prospect. Jeff Rothenberg (1998), for example, has been exploring a process by which electronic records would be "bundled" with the software on which they were written and with a terminal emulator as well, making the entire archive self-contained electronically and retrievable on any future generations of hardware. The emulator is a "program that mimics the behavior of the hardware" (p. 15). An example of this is the earlier form of e-mail, predating LAN systems, which was accessible through an institutional mainframe. To gain mainframe access, one had to enter a terminal type, such as "VT100," one of many terminal emulator protocols, to make the personal computer compatible with the mainframe and capable of "talking" to it. As Rothenberg points out in his bundled text/software/terminal emulator plan, this scheme has advantages over migration and standards, both of which are limited by the evolution of information technology.

As long as our culture sees the value of rare books, manuscripts, and other traditional resources for research, libraries and archives will continue to support scholarly research in the traditional way. Libraries will create digital surrogate copies of some of their collections, but it is infeasible to believe that the entire corpus of our libraries' research resources will be converted retrospectively (see Zeidberg, 1993a, 1993b). Archivists will have to manage what does get converted and what is being created digitally from the outset if they are to fulfill their responsibility to future generations of scholars needing access to research resources. Jeff Rothenberg (1998) sums up the present predicament best:

Beyond having obvious pragmatic value, the digital documents we are currently creating are the first generation of a radically new form of record-keeping. As such, they are likely to be viewed by our descendants as valuable artifacts from the dawn of the information age. Yet we are in imminent danger of losing them even as we create them. We must invest careful thought and significant effort if we are to preserve these documents for the future. If we are unwilling to make this investment, we risk substantial practical loss, as well as the condemnation of our progeny for thoughtlessly consigning to oblivion a unique historical legacy. (p. 17)

Archivists will need to work with writers to authenticate and preserve electronic documents and records if they are to have those resources last

as long as have our traditionally formatted materials. They should take care to leave a clear trail from creation to preservation so that scholars in the future who need to gain access to these resources will have the same opportunity to research them as they now enjoy with our culture's rarest manuscripts and books. The content after all is the message for scholars—not the medium.

## REFERENCES

- Daniell, D. (1994). *Let there be light: William Tyndale and the making of the English Bible*. London: The British Library.
- George III, autograph letter to Thomas Townshend, 19 November 1782. (1782). (Huntington Library: HM 25755).
- Grafton, A. (1997). *Commerce with the classics: Ancient books and Renaissance readers*. Ann Arbor: University of Michigan Press.
- Graham, P. S. (1998). New roles for special collections on the network. *College & Research Libraries*, 59(3), 234-235.
- Hobson, A. (1998). Was there an Aldine Bindery? In *Aldus Manutius and Renaissance culture: Essays in memory of Franklin D. Murphy*. Florence: Leo S. Olschki.
- Mann, T. (1993). *Library research models: A guide to classification, cataloging, and computers*. New York: Oxford University Press.
- Rothenberg, J. (1995). Ensuring the longevity of digital information. *Scientific American*, 272(1), 24-29. Revised presentation version, January 1998, p. 15.
- Zeidberg, D. S. (1993a). Setting the course: The role of special collections in the library. *Rare Books & Manuscripts Librarianship*, 8(2), 106-111.
- Zeidberg, D. S. (1993b). The future of special collections in emerging information delivery programs. *Journal of Library Administration*, 19(1), 67-82.